

Dynamic Resource Allocation and Pricing for Shared Radio Access Infrastructure

Özgür Umut Akgül^{*†}, Ilaria Malanchini[†], Vinay Suryaprakash[†], Antonio Capone^{*}

^{*}Dipartimento di Elettronica, Informazione e Bioingegneria Politecnico di Milano, Italy

Email: {oezguerumut.akguel, antonio.capone}@polimi.it

[†]Nokia Bell Labs, Stuttgart, Germany

Email: {ilaria.malanchini, vinay.suryaprakash}@nokia-bell-labs.com

Abstract—Flexible resource sharing at short time scales in multi-tenant shared radio access networks has proven to be quite a challenge. In this study, we develop a techno-economic model that enables dynamic short-term resource sharing as well as resource pricing, while simultaneously collecting revenue for network expansion. In order to regulate the resource costs and to prevent monopolization of resources, we define a unit cost of resources which can be scaled dynamically. The proposed framework allows operators to meet their individual utility targets while optimizing their expenditures based on their respective budgets. This work demonstrates that dynamic short timescale resource sharing can help network operators achieve their utility targets while minimizing their total expenditure.

I. INTRODUCTION

A multitude of applications, driven in part by the Industry 4.0 initiative, are envisioned for future networks (5G and beyond), [1]. Most of these applications require not only high data rates, but also low latencies. One of the potential solutions to this problem is considered to be denser and more heterogeneous network deployments, [2]. This, however, places an enormous strain on the already decreasing profitability of mobile operators, [3] and thereby, necessitates a change in their current business modus operandi. One of the solutions proposed to cope with increasing operational costs and decreasing profitability is *Infrastructure Sharing*, [4]. As the name suggests, this idea proposes that mobile network operators (MNOs) share a common infrastructure in order to reduce their capital and operational expenditure as well as to offer their customers better prices, a larger number of services, and a better quality of service.

As detailed in the Organization for Economic Co-operation and Development (OECD) report, [4], infrastructure sharing can be undertaken at various levels. One of the most comprehensive methods of sharing is where there are multiple mobile virtual network operators (MVNOs) who lease or rent the infrastructure from an infrastructure provider and such a type of sharing is the focus of this paper. In general, sharing takes place based on service level agreements (SLAs) between the parties who intend to share the infrastructure and it usually takes the form of contracts which are enforced over a long period of time. However, the OECD report, [4, Pg. 65], also states: “savings from active sharing are greater than for passive sharing as a higher proportion of costs are shared”. Active sharing implies sharing radio access network resources

including the spectrum. This type of sharing, however, quickly becomes infeasible if today’s (i.e. long term) SLAs are used. This is because the MVNOs will not have the ability to accommodate fluctuations in their traffic and might often find themselves in scenarios where they risk being unable to cater to their customers. Active sharing, therefore, requires a more dynamic sharing methodology which allows MVNOs to share and trade resources at much shorter timescales, i.e., in a few seconds or minutes. In order for such a system to work, viz. for it to be profitable for all the parties involved, each of them should have a good understanding of their own budgets, the implications of short-term fluctuations on them, and an accurate estimate of their traffic load. It, therefore, becomes imperative that each of the parties involved, be it MVNOs or infrastructure providers, have sound techno-economic models that are able to estimate aspects like resource allocation, the required network expansion, and their implications on resource pricing. As detailed in Subsection I-B, this is precisely where today’s models fall short and this is an aspect this paper tries to address.

A. Contributions

In the interest of facilitating the active sharing promoted by the OECD, we propose a techno-economic model that allows dynamic short term resource sharing as well as short term price negotiations between the MVNOs and the infrastructure provider. The contributions of this paper are as follows:

- It provides a *first step* towards a more comprehensive *techno-economic market model* for radio access networks.
- It proposes a *short time scale* dynamic trading model wherein: i) the cost of resources is market driven, and ii) the MVNOs trade resources based on their ability to satisfy customer demands as well as meet their respective budget constraints.

B. Related Work

Relatively speaking, technological models have garnered more attention only in the recent past. Works such as [5]–[9] estimate the performance and provide a comparison of networks wherein both physical and virtual sharing of capacity or spectrum occurs. These works, however, tend to be system or technology dependent (e.g., focusing solely on LTE, etc.). Lately, there have been attempts in papers such as [10] or

[11] to provide more generic resource sharing algorithms. The economic aspects salient to MVNOs are, however, not considered in the aforementioned works.

Economic models, on the other hand, have been used quite extensively to motivate the need for network sharing. There are numerous works such as [4] and the references therein, which focus on various aspects related to the costs of sharing specific network components. More specifically, papers such as [12] and [13] argue in favor of site sharing, radio access network sharing, and core network sharing as ways towards a sustainable business platform for the future. Other works such as [14] and [15] also explore the relationship between network costs and the extent to which networks are shared. Admittedly, there is an implicit link between the technological and economic aspects when varying degrees of network sharing are explored. These works, however, do not shed sufficient light on the technological implications (i.e., on the ability to satisfy customer demands) of economic decisions made.

Another aspect - overlooked in most works - is the fact that the models proposed still focus on long term SLAs, which do not provide the flexibility required to enable active network sharing. An added degree of flexibility in the SLAs is provided in [11], where a sharing model, which allows the parties to deviate from the constraints agreed upon in the SLA to a certain extent while abiding by the SLA's constraints on average. This idea forms the basis of our work in this paper.

II. SYSTEM MODEL

In our model, we considered two different types of stakeholders, i.e. a single infrastructure provider and multiple MVNOs. Let M represent the set of MVNOs and let $|M|$ be its cardinality. Then, let K represent the set of active users that are distributed between the MVNOs and let the set of active users of MVNO m be represented by K_m . We assume that the decisions taken by a base station's scheduler are not directly effected by schedulers in the neighbouring base stations and thereby, focus on the downlink of a single base station. Like [11], we assume that there exists an initial agreement between the infrastructure provider and the MVNOs, which sets the initial values of the network resources to be shared. However, unlike [11], we do not consider statically shared network resources; instead, MVNOs update their share of the network resources based on their respective traffic and utility targets.

A. Notations and the Model

To ensure consistency and clarity, this work uses the same notations as those used in [11]. In this framework, $S_m \in [0, 1]$ represents the sharing ratio, i.e., the percentage of resources, for operator m based on predefined SLAs and $\Delta_m \geq 0$ denotes the maximum deviation from S_m (when averaged over a certain time window). The average resources that a particular MVNO gets cannot deviate from S_m by more than Δ_m within W . Now, recall that our goal is to further a more dynamic resource trading environment in which the MVNOs are free to pursue their individual interests. With this objective in mind, in the proposed framework, S_m and Δ_m are MVNO

specific variables that can be re-negotiated periodically, where the period specified by a time window W is determined by the infrastructure provider. Note that, in this work, we consider the existence of just a single infrastructure provider, who is not subject to conventional market pressures.

Similar to most works having to do with schedulers, time is discretized and partitioned into time slots. As in [11], $x_k[n]$ denotes the fraction of resources assigned to the user k at time slot n . Depending on the resources assigned, the deviation of operator m from S_m at time slot n is denoted by $\epsilon_m[n]$. Additionally, $r_k[n]$ represents the achievable rate of user k during the time slot n . Apart from the notation used in [11], we also define new parameters relevant to a techno-economic model. In this model, each operator has a budget, B_m , that can be spent at any time instance n . We define three types of cost, namely: capital expenditure (CapEx), operational expenditure (OpEx), and pressure cost denoted by C_{ca} , C_{op} , and C_{pre}^m (for MVNO m), respectively. The MVNO has a CapEx proportional to its S_m , whereas the OpEx is based on the actual resources obtained. This definition incentivizes the MVNOs to utilize the added flexibility to deviate from the original resource sharing limits agreed upon by coupling each MVNO's expenses with their needs and budget constraints.

The pressure cost, C_{pre}^m , ensures that the costs of network resources scale according to their demand and, from an infrastructure provider's point of view, provides a steady revenue stream for expenses like (longterm) capacity expansion. In the market model considered, when the number of available resources decreases, the cost of purchasing a given unit of resource increases. In order to create this inversely proportional dependence between the scarcity of resources and their cost, the term $\xi_m[n]$ - reflecting the difference between an operator's utility target ($U_{th,m}$) and the actual utility they obtained - is used. Throughout the paper, we refer to resource scarcity as the case where $\xi_m > 0$ and $\sum_{k \in K} x_k = 1$ and to resource surplus as the case where $\xi_m = 0$, $\forall m \in M$ and $\sum_{k \in K} x_k[i] < 1$. The product of $\xi_m[n]$ and C_{pre}^m , therefore, provides the surcharge for the resources requested at time slot n . Since the pressure cost is proportional to a given operator's utility target, a long term aggregate of this cost results in the amount (proportional to the sum of the utility requirements of all MVNOs) which should be invested towards network or capacity expansion.

B. Assumptions

The salient assumptions are as follows:

- 1) *Operator's gap*, $\xi_m[n]$, gives complete information about the *additional resources* required to satisfy an MVNO's target.
- 2) All the traffic is *elastic*, i.e., the traffic is not sensitive to delays.

III. FORMULATION AND ANALYSIS OF THE MODEL

A. Problem Formulation

Based on the notation defined in Section II, the generic optimization problem solved at the base station's scheduler

PREPRINT VERSION

$$\min f(\xi_m[n], S_{\max}) \quad (1a)$$

$$\text{s.t.} \quad S_{\max} \geq \max(S_m, 1 - S_m), \quad \forall m \in M, \quad (1b)$$

$$\xi_m[n] \geq \max(0, U_{\text{th},m} - \frac{1}{(a+1)|K_m|} \sum_{i=n-a}^n \sum_{k \in K_m} U_k(x_k[i], r_k[i])), \quad \forall m \in M, a \equiv (n-1 \bmod W), \quad (1c)$$

$$\epsilon_m[n] = \left(\frac{1}{(a+1)} \sum_{i=n-a}^n \sum_{k \in K_m} x_k[i] \right) - S_m, \quad \forall m \in M, \quad (1d)$$

$$|\epsilon_m[n]| \leq \Delta_m, \quad \forall m \in M, \quad \forall n \in N, \quad (1e)$$

$$\sum_{i=n-a}^n (S_m(C_{\text{ca}} + C_{\text{op}}) + \epsilon_m[i]C_{\text{op}} + \min(\xi_m[n]C_{\text{pre}}^m, B_m)) \leq B_m(a+1), \quad \forall m \in M, \quad (1f)$$

$$0 \leq \Delta_m \leq \max(S_m, 1 - S_m), \quad \forall m \in M, \quad (1g)$$

$$\sum_{k \in K} x_k[n] \leq 1, \quad x_k[n] \geq 0, \quad \forall k \in K, \quad (1h)$$

$$\sum_{m \in M} S_m \leq 1, \quad S_m \geq 0, \quad \forall m \in M. \quad (1i)$$

to dynamically optimize the resource allocation and pricing is given by Equations (1a)-(1i). Ideally, the scheduler has the achievable rates of users for each MVNO and schedules using these rates. However, since the algorithm is proposed for a real time scheduling problem, the optimizer does not know the rates in the future. Therefore, the selection of efficient S_m and Δ_m is not trivial as the MVNOs have to predict their future needs extremely accurately. In order to solve this challenge, the optimization problem (1a)-(1i) is split into a two stage optimization problem denoted by P1 and P2. The individual objective functions used and their respective constraints are explained in Subsection III-B.

Our optimization problem considers a continuous objective function (1a), which depends on two parts. The first part minimizes the total gap between the individual MVNO's desired utility and their actual utility values at time slot n . In the second part, in order to guarantee a fair distribution of the CapEx among MVNOs, we minimize S_{\max} , which denotes the maximum between the SLA based resources available to an MVNO (S_m) and the remaining resources ($1 - S_m$) as defined in constraint (1b). The minimum of the right-hand side (RHS) of (1b) can be achieved if S_m is selected equal to $1 - S_m$. Based on this logic, the optimizer assigns $S_m = \frac{1}{|M|}$, $\forall m \in M$, if the budgets of all MVNOs permit it. Therefore, this constraint, i.e. (1b), ensures that fairness is achieved in terms of the initial sharing of resources. The MVNO can obtain additional resources (if available) by selecting a higher Δ_m value. The gap of operator m , $\xi_m[n]$, is constrained by (1c). The first term within the maximization function in (1c) prevents the gap from being lower than zero, and reflects the fact that the network (provided by the infrastructure provider) is able to handle traffic effectively enough that no expansion is necessary even in the long run. The second term in the maximization function, on the other

hand, computes the difference between the desired utility – denoted by $U_{\text{th},m}$ – of a given MVNO m and the utility function $U_k(x_k[n], r_k[n])$ measured at time n , where $x_k[n]$ and $r_k[n]$ are the percentage of resources and the rate assigned to a user k during a particular slot n , respectively. Constraint (1d) sets the value of $\epsilon_m[n]$, which is the instantaneous deviation from the agreed sharing ratio. The first term on the RHS of the equation is the average resources that MVNO m obtained from the beginning of the current time window to the current time slot n , whereas the second term is the SLA based sharing ratio. When $\xi_m[n]$ and $\epsilon_m[n]$ are calculated, each W is considered to be independent of the other.

Constraint (1e), in which $\epsilon_m[n]$ is computed using (1d), limits the maximum deviation between the agreed sharing ratio and the obtained resources from exceeding Δ_m . Constraint (1f) is the budget constraint, which ensures that the overall expenditure in a time window cannot exceed an MVNO's budget for that time window. The operator pays both CapEx and OpEx for the fixed resource shares agreed upon in the SLA, S_m , which is accounted for by the first term in the summation on the left-hand side (LHS) of the inequality. However, by choosing a higher deviation Δ_m , the operator has the ability to increase or decrease their expenditure in relation to the costs computed using the SLA. This aspect is taken into account by incorporating $\epsilon_m[n]$ in the second term on the LHS of the inequality (1f). If the MVNO receives fewer resources than S_m , as can be observed from (1d), the second term of (1f) becomes negative and decreases the total cost. The third term on the LHS of (1f) is the pressure cost that is designed in order to regulate the demands of individual MVNOs and to introduce a notion of 'supply and demand' economics to these short term resource acquisitions. From an infrastructure provider's perspective, it also acts as a means to collect the necessary revenue for network expansion. Since

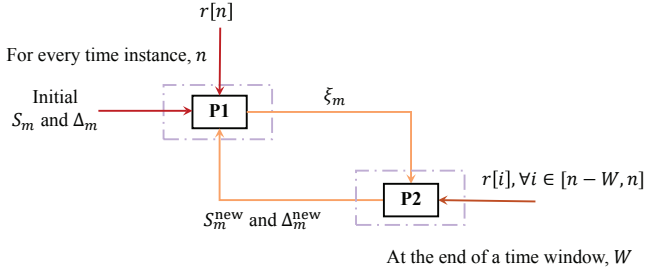


Fig. 1. Two stage optimization solved at the scheduler.

the MVNOs are not obligated to spend their entire budget during a given time slot, the unspent revenue from previous time slot can be used during the successive time slot. This effect is represented by the scaling factor $(1 + a)$ on the RHS of (1f), where $a \equiv (n - 1) \bmod W$. Then, constraint (1g) introduces the necessary coupling between Δ_m and S_m in order to prevent the MVNOs from trading resources that they do not have. Constraint (1h) ensures that the total number of resources consumed is always less than or equal to the total number of resources available and the network resources allocated to any user k cannot be lower than zero. Finally, constraint (1i) ensures non-negativity of the resources initially agreed upon in the SLA and it also prevents the sum of S_m over all the operators from being greater than one.

Note that the problem is presented in a non-linear form to improve readability. It can, nevertheless, be linearized with standard techniques.

B. Applied Algorithm

Owing to the many fluctuations encountered during wireless communications, the optimization problem to determine the most cost effective resource allocation for a given MVNO is solved in two steps denoted by P1 and P2 (detailed below) and as illustrated in Fig. 1. During the first step indicated by P1, the optimizer accepts S_m and Δ_m as input parameters and finds the optimum resource allocations that minimizes the total gap between each MVNO's target utility and the utility they achieved. During each time slot within a given time window W , the optimizer runs this resource allocation optimization (i.e., P1) using the respective rate estimates of the active users. At the end of W , the optimizer switches to the second step, i.e. P2.

$$\begin{aligned} \text{P1} &:= \begin{cases} (1a) & \min_{\xi_m, x_k, \epsilon_m} \sum_{m \in M} \xi_m[n] \\ \text{s.t.} & (1c)(1d)(1e)(1f)(1h) \end{cases} \\ \text{P2} &:= \begin{cases} (1a) & \min_{\xi_m, x_k, S_m, \Delta_m, \epsilon_m} \sum_{m \in M} \xi_m[n] + S_{\max} \\ \text{s.t.} & (1b) - (1i) \end{cases} \end{aligned}$$

During P2, the optimizer determines the optimal resource allocation for the previous time window (i.e., the window that just ended) using the knowledge of all the rates actually

achieved. Then, based on these 'optimum' resource allocations, the optimizer determines the optimal S_m and Δ_m , and updates their values for the upcoming time window. The update process is performed according to

$$S_m^{\text{new}} = (1 - \alpha_m)S_m^{\text{old}} + \alpha_m S_m^{\text{opt}}, \quad (2)$$

$$\Delta_m^{\text{new}} = (1 - \alpha_m)\Delta_m^{\text{old}} + \alpha_m \Delta_m^{\text{opt}}, \quad (3)$$

where α_m is the feature scaling coefficient and $S_m^{\text{opt}}, \Delta_m^{\text{opt}}$ are the optimum S_m, Δ_m values for the previous time window.

Both for P1 and P2, the optimizer uses the same objective function. However, since the variables of the two problems are different, the constraints that are applicable for each of the problems will also be different.

C. Effects of Feature Scaling

The input parameters for P1 during the upcoming time window are selected based on their initial values and the optimum values during the previous time window. However, the determination of the scaling coefficient is a challenging task as large values of α_m effectively leads to a memoryless network resource optimization and very small values result in a static network resource optimization. A comparison between different scaling coefficients is presented in Fig. 2. As presented in (4), the relative distance to the optimal (RDO) gives an understanding of how close the selected parameters are to their optimum values, ξ_m^{opt} . Note that due to (1c), the gap $\xi_m[n]$ cannot be negative for any time slot and $\xi_m[n] \geq \xi_m^{\text{opt}}[n], \forall n \in N$. For the special case of $\xi_m[n] = \xi_m^{\text{opt}}[n] = 0$, the RDO is assumed to be 0; therefore, $\text{RDO} \in [0, 1]$.

$$\text{RDO} = \frac{1}{|M|} \sum_{m \in M} \frac{\sum_{i=1}^N \xi_m[i] - \xi_m^{\text{opt}}[i]}{\sum_{i=1}^N \xi_m[i]}. \quad (4)$$

The dynamic scaling coefficient is presented in (5), where $\xi_m^{\text{opt}}[n]$ is the optimum gap calculated by the optimizer during P2. Since it is determined by the actual rates achieved, it gives an idea of the minimum achievable gap if the scheduler has a-priori knowledge of the rates.

$$\alpha_m = \frac{\left| \sum_{i=n-a}^n \xi_m[i] - \sum_{i=n-a}^n \xi_m^{\text{opt}}[i] \right|}{\sum_{i=n-a}^n \xi_m[i] + \sum_{i=n-a}^n \xi_m^{\text{opt}}[i]}, a \equiv (n - 1 \bmod W). \quad (5)$$

The scaling coefficient (α_m) provides information about the difference between the observed gap and the minimum gap achievable per time window. Therefore, α_m measures how close the real-time scheduler performs to the optimum. Since the network parameters are selected according to a given MVNO's targets, each of them has a different α_m parameter that reflects the optimality of their decision.

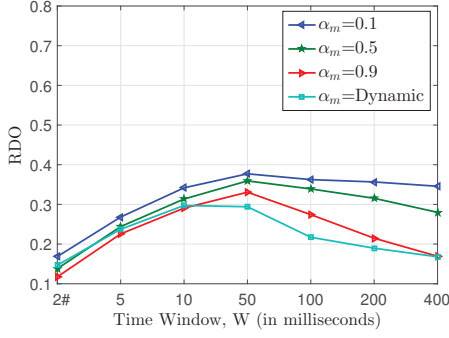


Fig. 2. Comparison of different scaling coefficients.

Fig. 2 presents the variation in the achievable gap for time windows of different lengths for various scaling coefficients. Since we would like to evaluate the merits of dynamic short time-scale resource sharing, we focus on time windows of length 50 – 200 ms. For this range, dynamic scaling is better than static scaling because it is better at coping with changes in window size. Therefore, the dynamic scaling coefficient is utilized in the simulations.

D. Effects of Pressure Cost

As previously mentioned, the motivation behind the introduction of the pressure cost is twofold. First, it helps regulate the price of resources (in scenarios of both resource surplus as well as resource scarcity), while ensuring that the price a given MVNO pays is proportional to their respective budget. Scaled pressure costs ensure that MVNOs will have the same chance to obtain resources and will be charged in proportion to their budgets. More specifically, the operators will not face scenarios where neither the purchase of resources nor the pressure costs are unaffordable. Therefore, the pressure cost is defined as

$$C_{\text{pre}}^m = \frac{B_m}{\sum_{m \in M} B_m} \times C_{\text{pre}}^{\text{unit}}, \quad (6)$$

where $C_{\text{pre}}^{\text{unit}}$ is the unit pressure cost of a given resource.

Second, in cases where the gap between the desired utility of MVNOs and their achieved utility is non-zero, since the pressure cost is proportional to the difference between the actual and desired utility values for each MVNO, the infrastructure provider – by means of the aggregated pressure costs collected – has the added advantage of knowing exactly how much has to be invested in capacity expansion.

IV. SIMULATION RESULTS

In this section, the simulation results of the two-step optimization problem are detailed.

A. Parameters and the Scenarios Studied

In order to analyze the applicability of our mathematical model, we considered the downlink of a base station that is shared by three MVNOs, i.e. $|M| = 3$. All the users are uniformly distributed throughout the coverage area of the base

TABLE I
THE APPLIED PARAMETERS AND THEIR VALUES.

Parameter	Definition	Value
C_{ca}	CapEx Cost	35.97
C_{op}	OpEx Cost	25.69
U_{th}^m	Rate target of operator m	2 bps/Hz
$ K_m $	Cardinality of the set of active users	1
W	Time window	100 ms
$ M $	Number of MVNOs	3
N	Duration of simulation	5000 ms
B_1	Budget of MVNO 1	88.45
B_2	Budget of MVNO 2	100
B_3	Budget of MVNO 3	56.17
$C_{\text{pre}}^{\text{unit}}$	Unit pressure cost per resource	35.97

station and, at each time slot, only one user from each operator becomes active. The simulation is run on a standard commercially available laptop for $N = 5000$ time slots, where each time slot is assumed to be 1 ms long, and the total run time of the algorithm (including both P1 and P2) is 0.998 sec. All the costs as well as the budgets are normalized to take values between 0 and 100 ($C_{\text{ca}}, C_{\text{op}}, C_{\text{pre}}^m, B_m \in [0, 100], \forall m \in M$) such that they can be considered as a generic value which can be spent during each time slot n . It is important to mention that the values of the budgets and costs are purely illustrative, whose purpose is to help understand the characteristic behavior of the model. Since the actual values that MVNOs use will merely be affine functions of the values used here, the behavior observed remains unchanged.

We model the channel between the user and the base station using a frequency-flat block fading channel with i.i.d. Rayleigh coefficients – resulting in exponentially distributed random channel gains $|h_k[n]|^2$. The Signal to Interference-plus-Noise Ratio (SINR) at any time instance is calculated as

$$\gamma_k[n] = |h_k[n]|^2 \text{SINR}_k, \quad (7)$$

where SINR_k is the average SINR of user k . This is calculated according to the Okumura-Hata propagation model as $\text{SINR}_k = P d_k^{-\alpha} / (\sigma^2 + I_0)$, where P is the transmit power (in Watts [W]), d_k is the user's distance to the base station (in meters [m]), α is the path-loss exponent, σ^2 is the thermal noise, and I_0 is the average interference power. Based on this, the spectral efficiency of a user (in bits/s/Hz) at time n is

$$r_k[n] = \log_2(1 + \gamma_k[n]). \quad (8)$$

Although the utility function can be something more intricate, the utility of the operator is measured in terms of the actual rate that a given MVNO's user achieves. Therefore, in a user-centric manner, the utility of each user is measured as

$$U_k(x_k[n], r_k[n]) = x_k[n] r_k[n]. \quad (9)$$

B. Performance Results

In Fig. 3, we begin by comparing the ability of various scaling coefficients to cope with variations in ξ_m due to the updates in the values of S_m and Δ_m caused by using (2) and

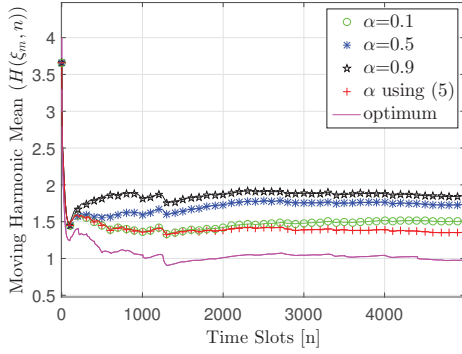


Fig. 3. Moving harmonic mean of the ‘total’ gap of the MVNOs computed over the all the previous time slots up to the current time slot n , for a time window $W = 100$ ms.

(3). The results of using various scaling coefficients are also compared with the case where the achievable rates for the upcoming time slots are known and the optimization problem can be solved for the entire time window as a whole. Fig. 3 plots the variations in the moving harmonic mean of ξ_m , $H(\xi_m, n)$, over all the time slots up to the time slot n in order to ensure that its ‘peak’ variations are more accurately captured than can be done when the arithmetic mean is used. We observe that the dynamic scaling coefficient computed using (5) outperforms the fixed scaling coefficients and is, as a result, closest to the ‘optimal’ case, i.e., the case where the rates for the subsequent time slots are known.

Fig. 4 and Fig. 5 document the variations in S_m and Δ_m , under two different cost scenarios. The objective function minimizing the maximum S_m in (1a), results in the same value of S_m for all the MVNOs as long as the MVNOs have the necessary budgets. Therefore, for the sufficient budget scenario (listed in Table I), since all the MVNOs never face a budget shortfall, $S_m = 0.33$ for all the MVNOs (Fig. 4(a)). Since each MVNO has the same S_m and incurs the same CapEx, this can be considered as the cost incurred to enter the coalition. In contrast to this initial sufficient budget scenario, for the second scenario (Fig. 4(b)) all the costs are doubled and obtaining network resources becomes more expensive. For this case, due to the infeasibility of the MVNOs’ budgets, S_m takes a smaller value than 0.33. By decreasing S_m , MVNOs decrease their overall CapEx and also try to achieve their objective in (1a). However, this CapEx adjustment is required only during budget shortfall in order to avoid underutilized resources. Fig. 5(a) shows the changes in Δ_m when the MVNOs have sufficient budgets. Since the window W is large, MVNOs have the ability to balance their utility targets and resource consumption based solely on S_m and their willingness to trade short-term resources, i.e. Δ_m , decreases over time. However, during a budget shortfall, the MVNOs are not able to buy enough resources due to budget infeasibility. Therefore, they have a higher incentive to share unused resources. In other words, as observed in Fig. 5(b), the MVNOs compensate for fluctuations in their resource requirements using Δ_m .

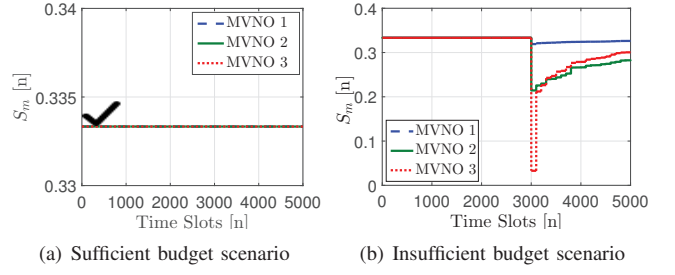


Fig. 4. Variation in S_m over time ($W = 100$ ms).

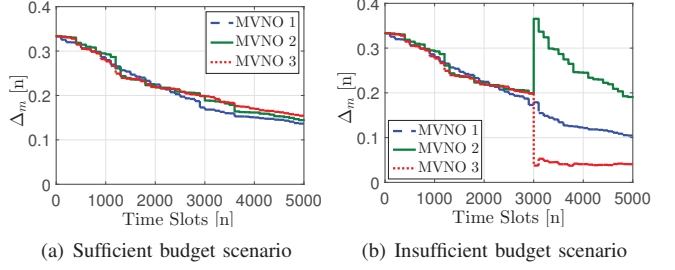


Fig. 5. Variation in Δ_m over time ($W = 100$ ms).

Fig. 6 and Fig. 7 provide additional insights into fairness in terms of resource distribution when there is no budget shortage. Fig. 6 is the cumulative distribution function (CDF) of the achieved rates for each MVNO. For all the MVNOs, we see that the probability of having a rate equal to 0 Mbps is around 0.6, which is a direct result of insufficient resources and maximum-rate scheduling. More specifically, the scheduler assigns all the resources to the user with the best channel conditions. Therefore, in a crowded network with similar user distributions, each MVNO can have the channel for $1/|M|$ of the time. Fig. 6 also shows that the MVNOs’ achievable rate distributions are very close to each other. This similarity shows that, despite the initial differences in resource distributions, MVNOs achieve similar rates on a relatively long-term. Fig. 7 plots the fluctuations in the moving arithmetic mean of ξ_m , $A(\xi_m, n)$, over all the time slots up to the time slot n . Despite the large deviations due to the channel quality and the initial S_m and Δ_m values, it is seen that $A(\xi_m, n)$ attains a stable characteristic around 2000 ms. The fluctuations observed till 2000 ms can be attributed to the transient state of the model and the non-optimal selection of S_m and Δ_m . However, after this point, they reach a steady state which suggests that no further improvements can be achieved just by changing S_m and Δ_m for given channel conditions.

Finally, the costs for sharing over various time windows and for not sharing are given in Fig. 8 for each MVNO. For the no-sharing scenario, the MVNOs are assumed to have their own infrastructure; whereas, for the static sharing case, the MVNOs share a fixed portion of resources. For static sharing case, since the MVNOs share a fixed portion of the resources, we assume that they will also share the overall expansion cost equally. The MVNOs’ costs are calculated using (1f) and averaged over the simulation duration N . As observed in

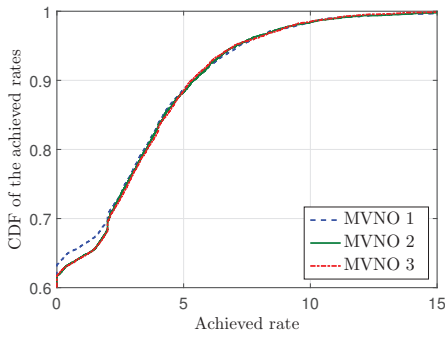


Fig. 6. Empirical CDF of achieved rates.

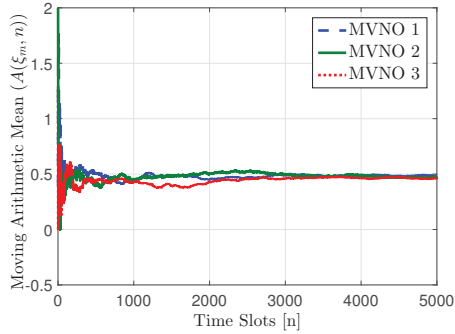


Fig. 7. Moving arithmetic mean of the ‘total’ gap of the MVNOs computed over all the previous time slots up to the current time slot n , for $W = 100$ ms.

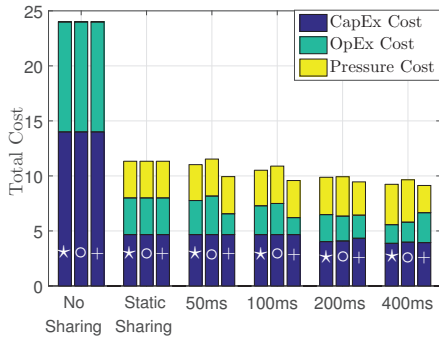


Fig. 8. Distribution of the costs (i.e. CapEx, OpEx and pressure) over MVNOs for different sizes of time windows and sharing options. (In the figure, \star represents MVNO 1, \circ represents MVNO 2 and, $+$ represents MVNO 3.)

Fig. 8, our framework provides an expenditure scaling based on the MVNOs’ utility targets and the resources they utilize. In conclusion, by using a more flexible model, we obtain a higher spectral efficiency than in static/no sharing scenarios (as shown in [11]) at comparable costs while ensuring that the MVNOs pay solely for what they use.

V. CONCLUSION

In this paper, we proposed a novel dynamic pricing and resource sharing algorithm for multi-tenant networks. The framework proposes a real-time wireless resource market and adjusts resource prices based on their scarcity and the need

for possible expansion in the future. This models also enables MVNOs to adjust their total expenditure based on their utility targets and the flexibility that can be tolerated while achieving them. It also imposes fairness in terms of the MVNO’s SLA based sharing ratio, which is considered as the cost of entering the coalition. This model affords the MVNOs the ability to adjust their total expenditure based on their individual budgets and (user-dependent) utility targets. Finally, by proposing pressure costs proportional to the MVNOs’ budgets, the threat of monopolization is reduced by ensuring that MVNOs with large budgets are penalized if they try to hoard resources in order to artificially inflate the unit cost of resources.

ACKNOWLEDGEMENT

The authors would like to thank Dr. -Ing. Thomas Haustein for his valuable suggestions which gave rise to this work.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 643002.

REFERENCES

- [1] NGMN, “5G White Paper,” 2015. [Online]. Available: <http://www.ngmn.de/5gwhite-paper.html>
- [2] Alcatel-Lucent, “5G is coming: Are you prepared?” 2015. [Online]. Available: <http://www2.alcatel-lucent.com/landing/5g/>
- [3] Alcatel-Lucent Bell Labs, “The declining profitability trend of mobile data: What can be done?” 2011. [Online]. Available: http://www3.alcatel-lucent.com/belllabs/advisory-services/documents/Declining_Profitability_Trend_of_Mobile_Data_EN_Market_Analysis.pdf
- [4] OECD, “Wireless market structures and network sharing,” 2014. [Online]. Available: <http://dx.doi.org/10.1787/5jxt46dzl9r2-en>
- [5] A. P. Avramova and V. B. Iversen, “Radio access sharing strategies for multiple operators in cellular networks,” in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 1113–1118.
- [6] V. Heinonen, P. Pirinen, and J. Iinatti, “Capacity gains through inter-operator resource sharing in a cellular network,” in *11th International Symposium on Wireless Personal and Multimedia Communications (WPMC)*, Sept 2008.
- [7] Y.-T. Lin, H. Tembine, and K.-C. Chen, “Inter-operator spectrum sharing in future cellular systems,” in *Global Communications Conference (GLOBECOM)*, 2012 IEEE, Dec 2012, pp. 2597–2602.
- [8] J. S. Panchal, R. Yates, and M. M. Buddhikot, “Mobile network resource sharing options: Performance comparisons,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4470–4482, 2013.
- [9] G. Salami and R. Tafazolli, “Interoperator dynamic spectrum sharing (analysis, costs and implications),” *International Journal of Computer Networks (IJCN)*, vol. 2, pp. 47–61, 2010.
- [10] J. Luo, J. Eichinger, Z. Zhao, and E. Schulz, “Multi-carrier waveform based flexible inter-operator spectrum sharing for 5G systems,” in *Dynamic Spectrum Access Networks (DYSPAN)*, 2014 IEEE International Symposium on, April 2014, pp. 449–457.
- [11] I. Malanchini, S. Valentin, and O. Aydin, “Wireless resource sharing for multiple operators: Generalization, fairness, and the value of prediction,” *Computer Networks*, vol. 100, pp. 110 – 123, 2016.
- [12] F. Berkers, G. Hendrix, I. Chatzicharistou, T. De Haas, and D. Hamera, “To share or not to share?” in *Intelligence in Next Generation Networks (ICIN)*, 2010 14th International Conference on, Oct 2010.
- [13] D.-E. Meddour, T. Rasheed, and Y. Gourhant, “On the role of infrastructure sharing for mobile network operators in emerging markets,” *Computer Networks*, vol. 55, no. 7, pp. 1576–1591, 2011.
- [14] I. Malanchini and M. Gruber, “How operators can differentiate through policies when sharing small cells,” in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015.
- [15] Y.-K. Song, H. Zo, and S. Lee, “Analyzing the economic effect of mobile network sharing in korea,” *ETRI Journal*, vol. 34, no. 3, pp. 308–318, 2012.